

## Systematische Gen-Suche in der Incyte LifeSeq Datenbank

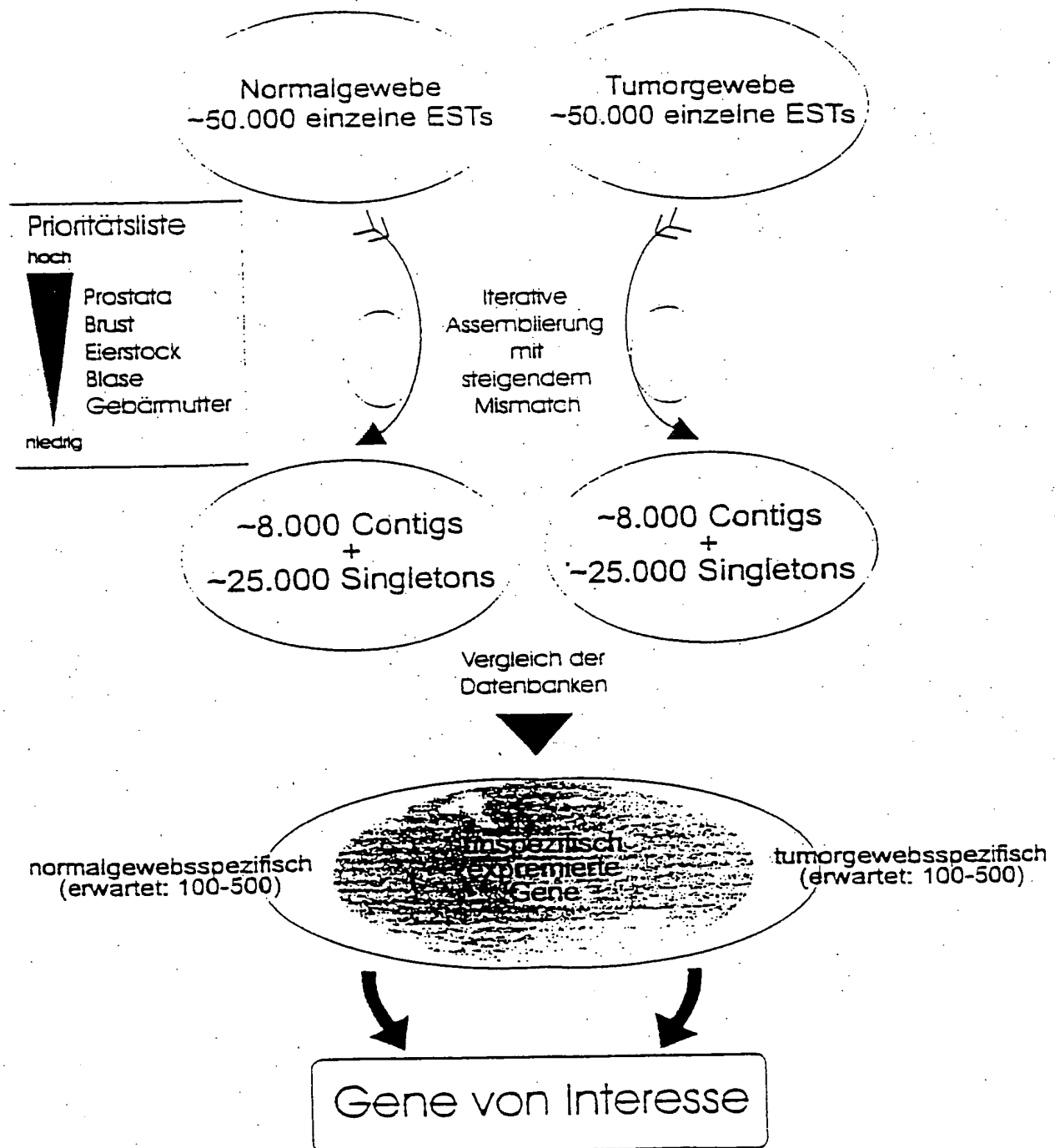


Fig. 1

Systematic Gene Search in the Incyte LifeSeq Database

Normal tissue  
~50,000 individual ESTs

Tumor tissue  
~50,000 individual ESTs

Priority list  
High

Prostate  
Breast  
Ovary  
Bladder  
Uterus

Iterative assembling  
with  
increasing mismatch

Low

~8,000 contigs  
+  
~25,000 singletons

~8,000 contigs  
+  
~25,000 singletons

Comparison of databases

normal tissue-  
specific  
(expected: 100-500)

nonspecifically  
expressed genes

tumor tissue-  
specific  
(expected: 100-500)

Genes of Interest

Figure 1

2/10

## Principle of EST Assembly

~50,000 ESTs per tissue

Assembly at 0% mismatch  
with GAP4 (Staden)

Contigs

Singletons

Contigs increasing in  
number and lengthIterative assembly with  
increasing mismatch  
(1%, 2%, 4%)

5000-6000 contigs

~25,000 other singletons

~30,000 consensus-  
sequences per tissue

Figure 2a

# Prinzip der EST-Assemblierung

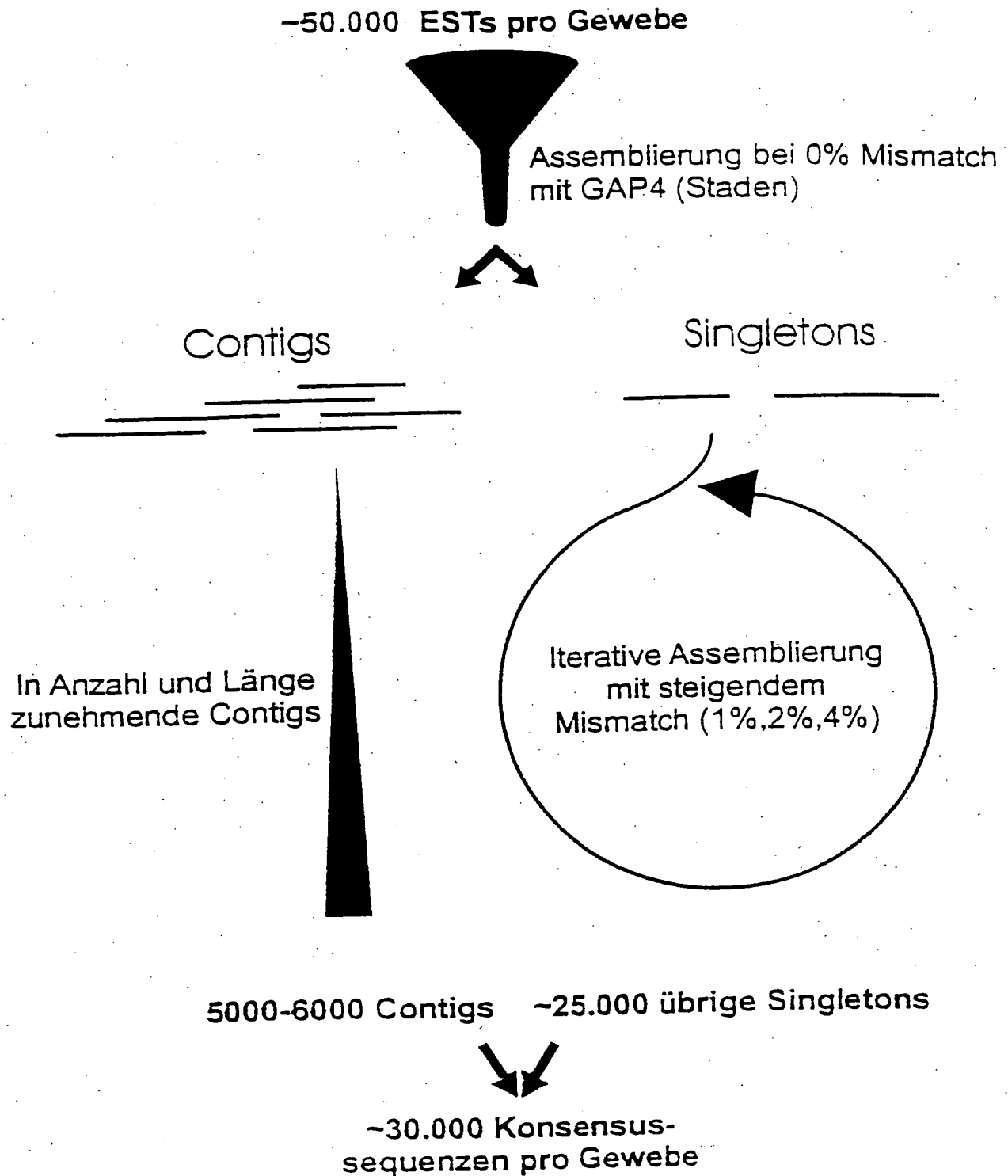


Fig. 2a

3/10

~50,000 ESTs of a tissue (e.g.: uterus tumor)

GAP4 Assembly 1st Round:  
minimum initial match: 20  
maximum pads per read: 8  
maximum percent mismatch: 0

GAP4 Database 1	unassembled
Contigs 1    Singletons 1	ESTs

GAP4 Assembly 2nd Round:  
minimum initial match: 20  
maximum pads per read: 8  
maximum percent mismatch: 1

GAP4 Database 2	unassembled
Contigs 2    Singletons 2	ESTs

GAP4 Assembly 3rd Round:  
minimum initial match: 20  
maximum pads per read: 8  
maximum percent mismatch: 2

GAP4 Database 3:	unassembled
Contigs 3    Singletons 3	ESTs

Figure 2b1

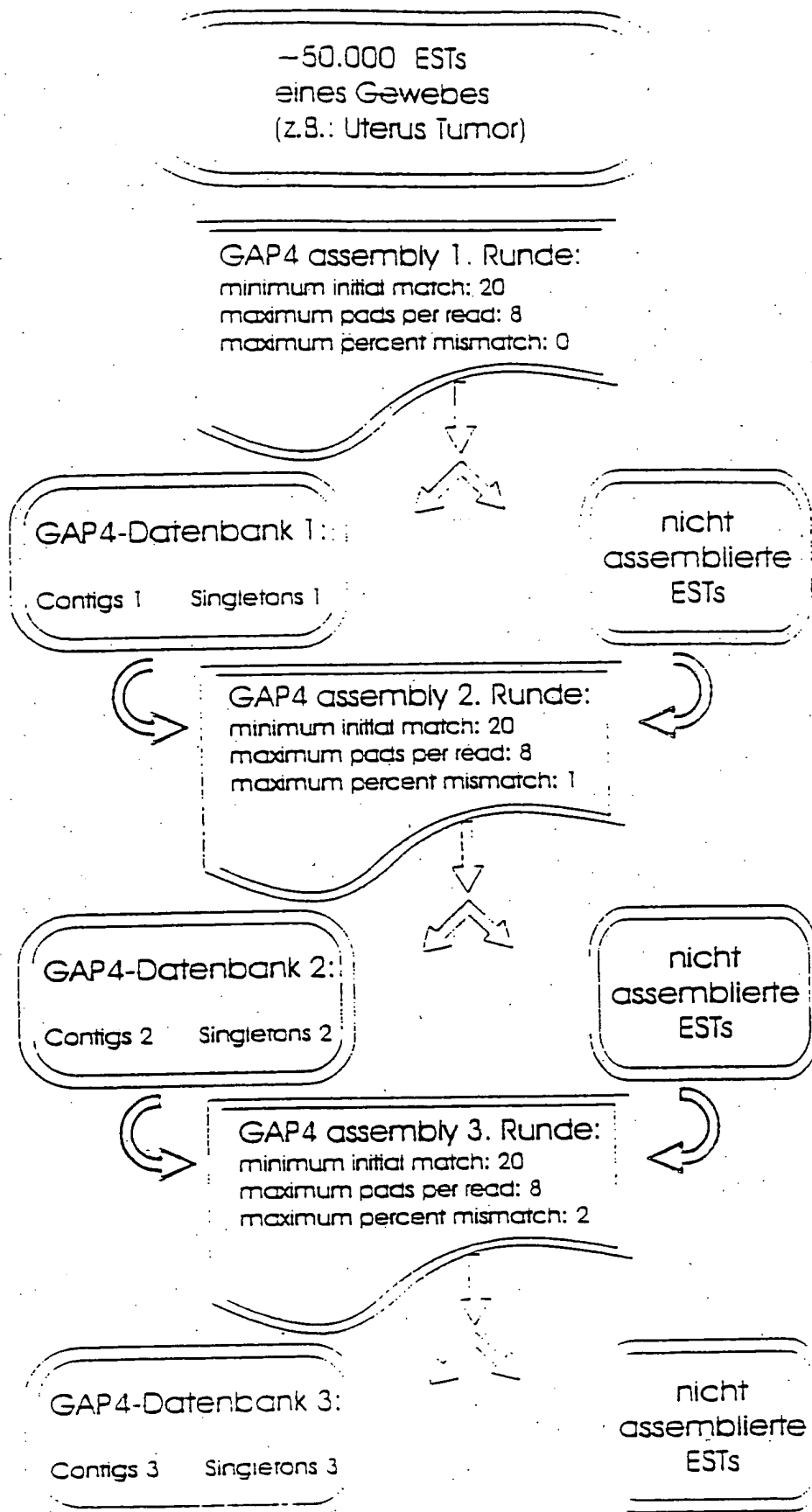


Fig. 2b1

4/10

GAP4 Database 3:  
Contigs 3    Singletons 3

unassembled  
ESTs

Consensus 3

GAP4 Assembly 4th Round:  
minimum initial match: 20  
maximum pads per read: 8  
maximum percent mismatch: 2

GAP4 Database 4:  
Contigs 4    Singletons 4

unassembled  
ESTs

Consensus 4

GAP4 Assembly 5th Round:  
minimum initial match: 20  
maximum pads per read: 8  
maximum percent mismatch: 4

GAP4 Database 5:  
Contigs 5    Singletons 5

unassembled  
ESTs 5

Consensus 5

Singletons 5

Figure 2b2

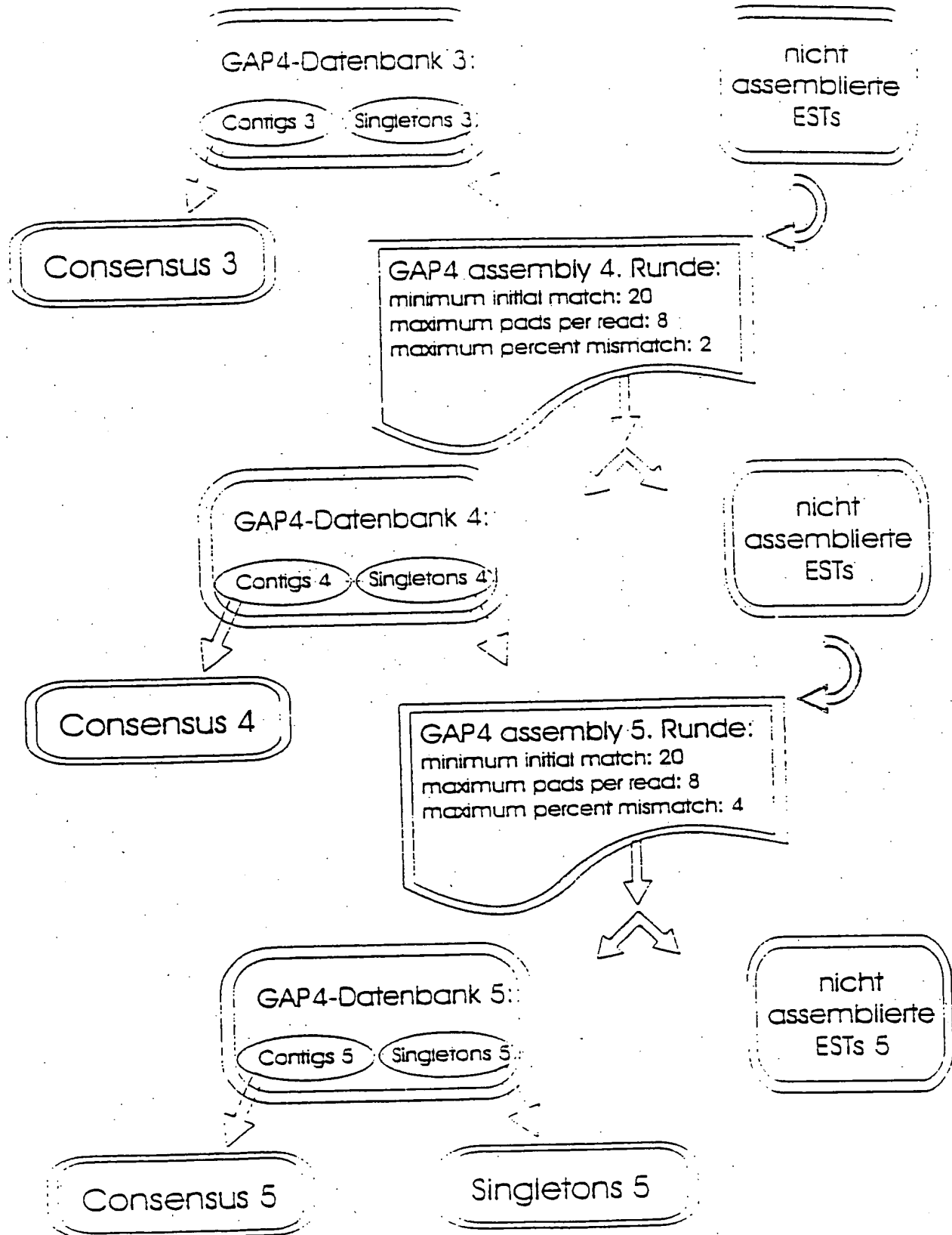


Fig. 2b2



5/10

Consensus 3

Singletons 5

Consensus 4

unassembled  
ESTs 5

Consensus 5

GAP4 Assembly 6th Round:  
minimum initial match: 20  
maximum pads per read: 8  
maximum percent mismatch: 4

Assembled database  
of a specific tissue  
(e.g.: uterus tumor)

Figure 2b3

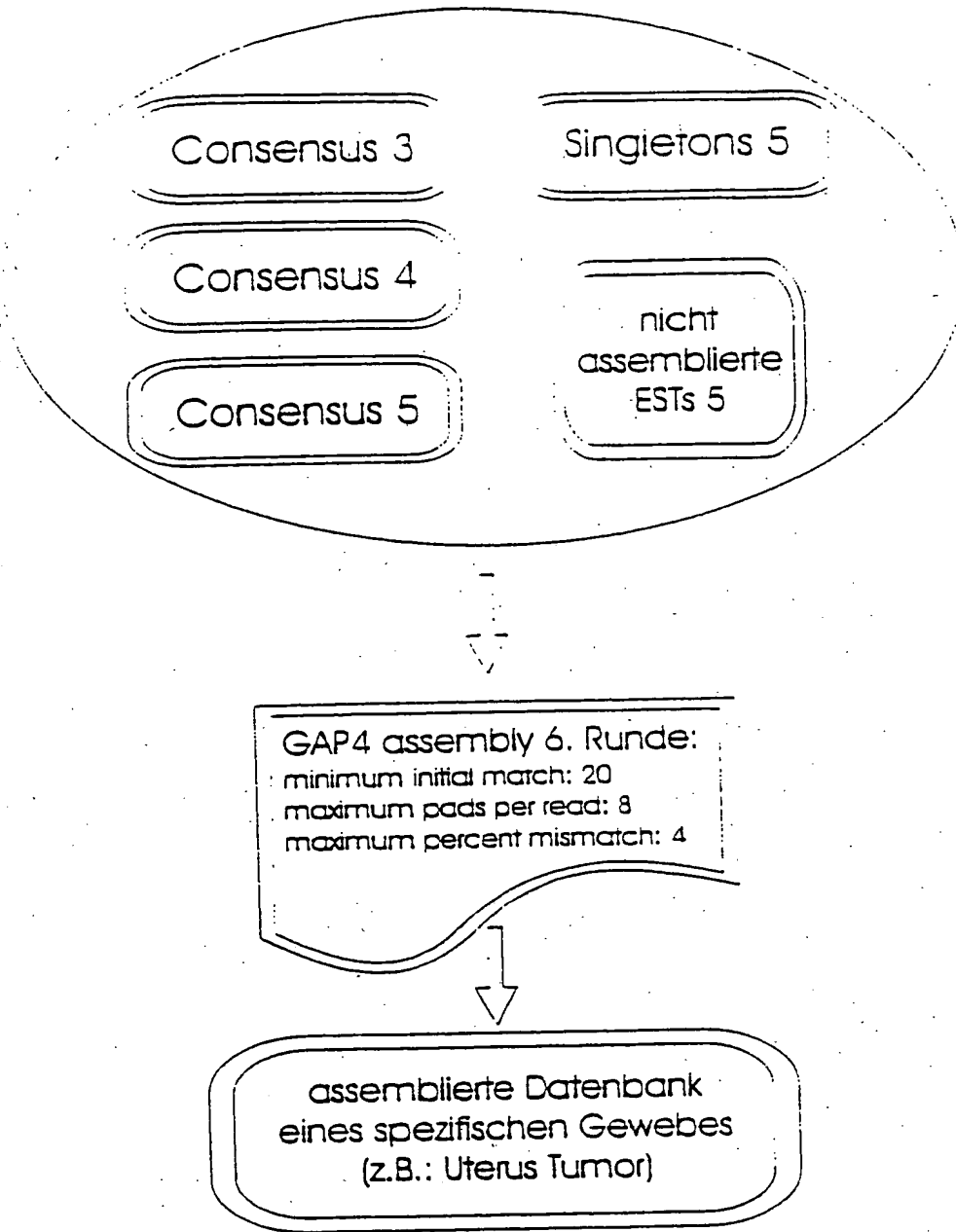


Fig. 2b3

6/10

Assembled database  
of a specific tissue  
(e.g.: uterus tumor)

Consensus 6

Read-in as singletons

Database  
of a specific tissue  
(e.g.: uterus tumor)

Database of a second  
specific tissue  
(e.g.: normal uterus)

GAP4 Assembly  
minimum initial match: 20  
maximum pads per read: 8  
maximum percent mismatch: 4

Tumor tissue-  
specific ESTs

Non-tissue-  
specific ESTs

Normal tissue-  
specific ESTs

Fig. 2b4

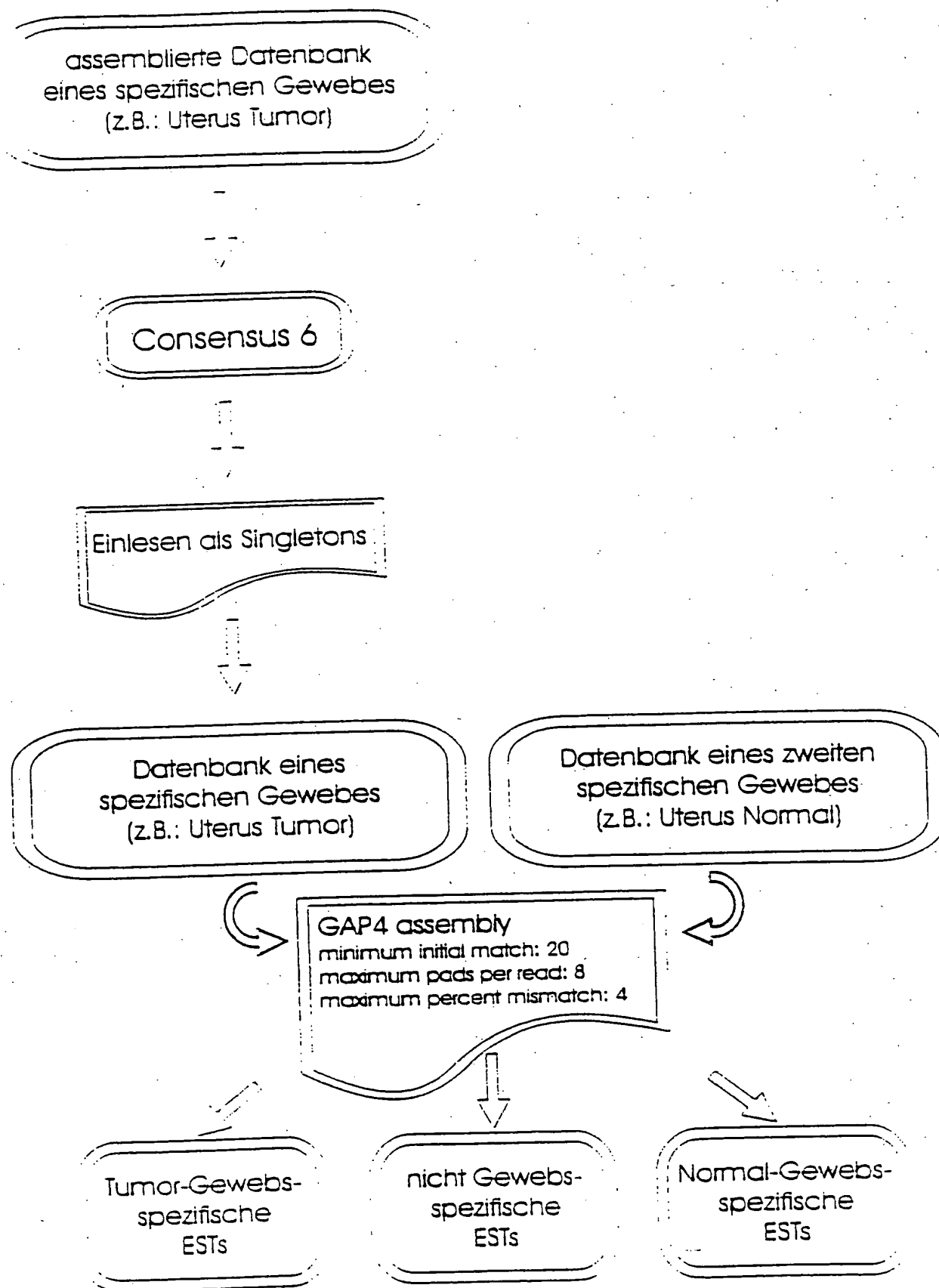


Fig. 2b4

7/10

In silico subtraction of gene expression in various tissues

~30,000 consensus sequences  
normal tissue

~30,000 consensus sequences  
cancer tissue

Assembly at 4% mismatch

Normal tissue  
Specific genes

Cancer tissue  
Specific genes

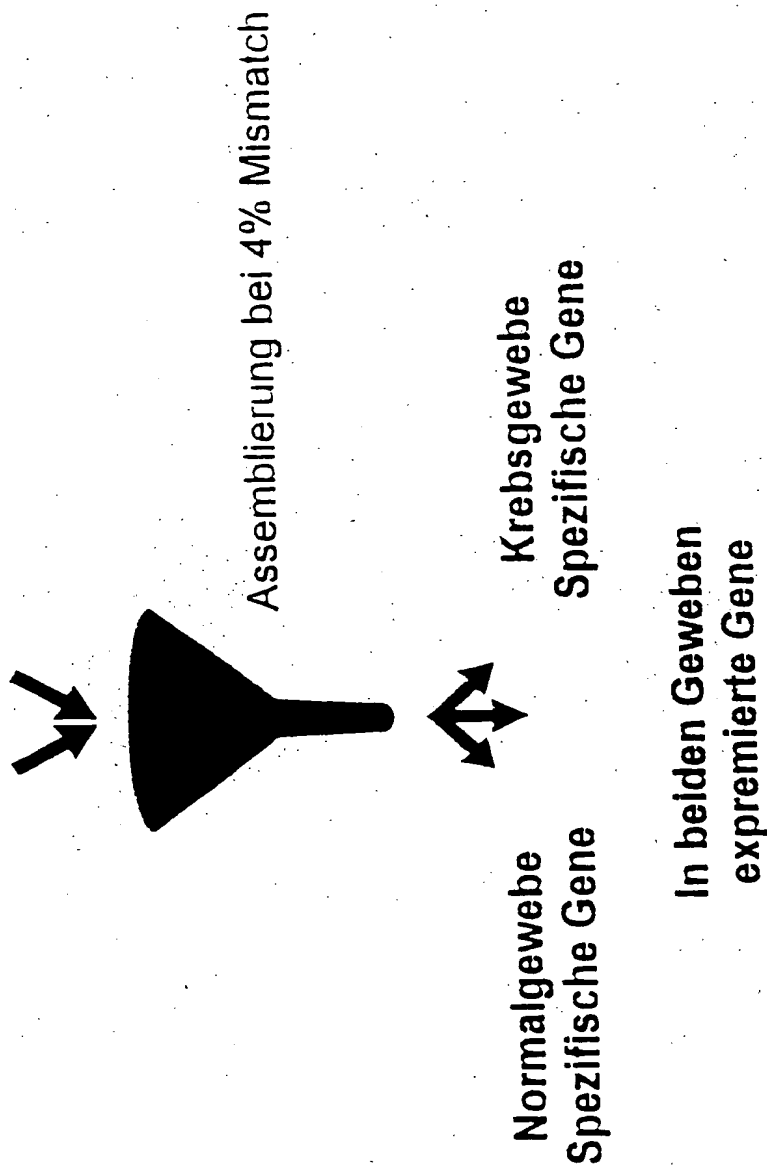
Genes expressed in both tissues

Figure 3

# In silico Subtraktion der Genexpression in verschiedenen Geweben

~30.000 Konsensussequenzen  
Normalgewebe

~30.000 Konsensussequenzen  
Krebsgewebe



metaGen  
Genexpression in verschiedenen Geweben

Fig. 3

8/10

Genes of interest

Determination of tissue-specific expression  
via electronic Northern (INCYTE LifeSeq and  
public EST databases)

Candidate genes for tumor suppressors or  
tumor activators

Figure 4a

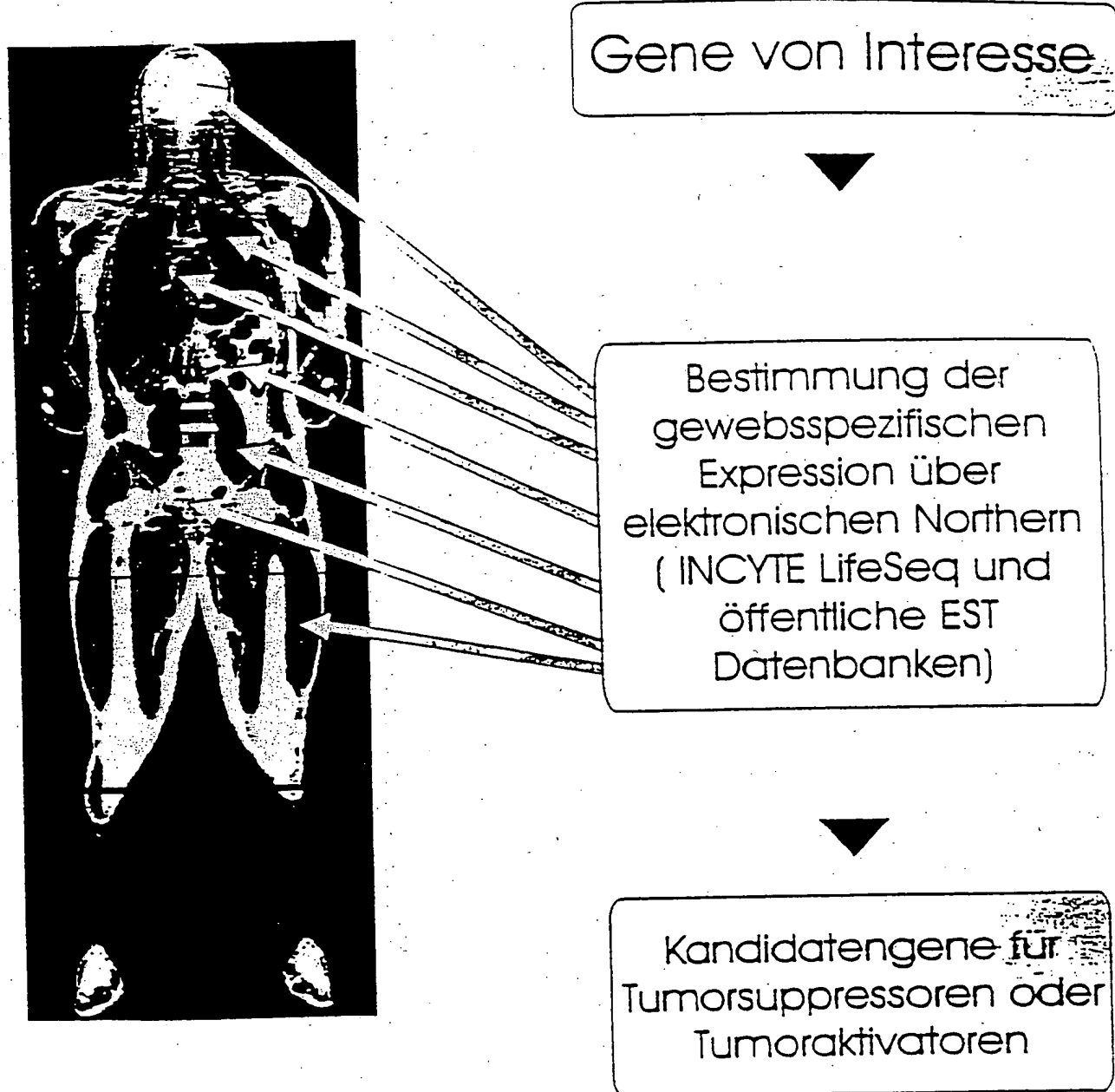


Fig. 4a



9/10

Partial cDNA sequence  
e.g., EST or contig  
S

...GCCTCAAGTTATC...

WHILE  $C_i > C_{i-1}$ 

Electronic Northern Blot

Fisher's Exact Test IF  $H_0$  EXIT

Automatic Lengthening

Consensus sequence C

...ATGTCCTAGCCTCAAGTTATCAGATGCAA...

Figure 4b

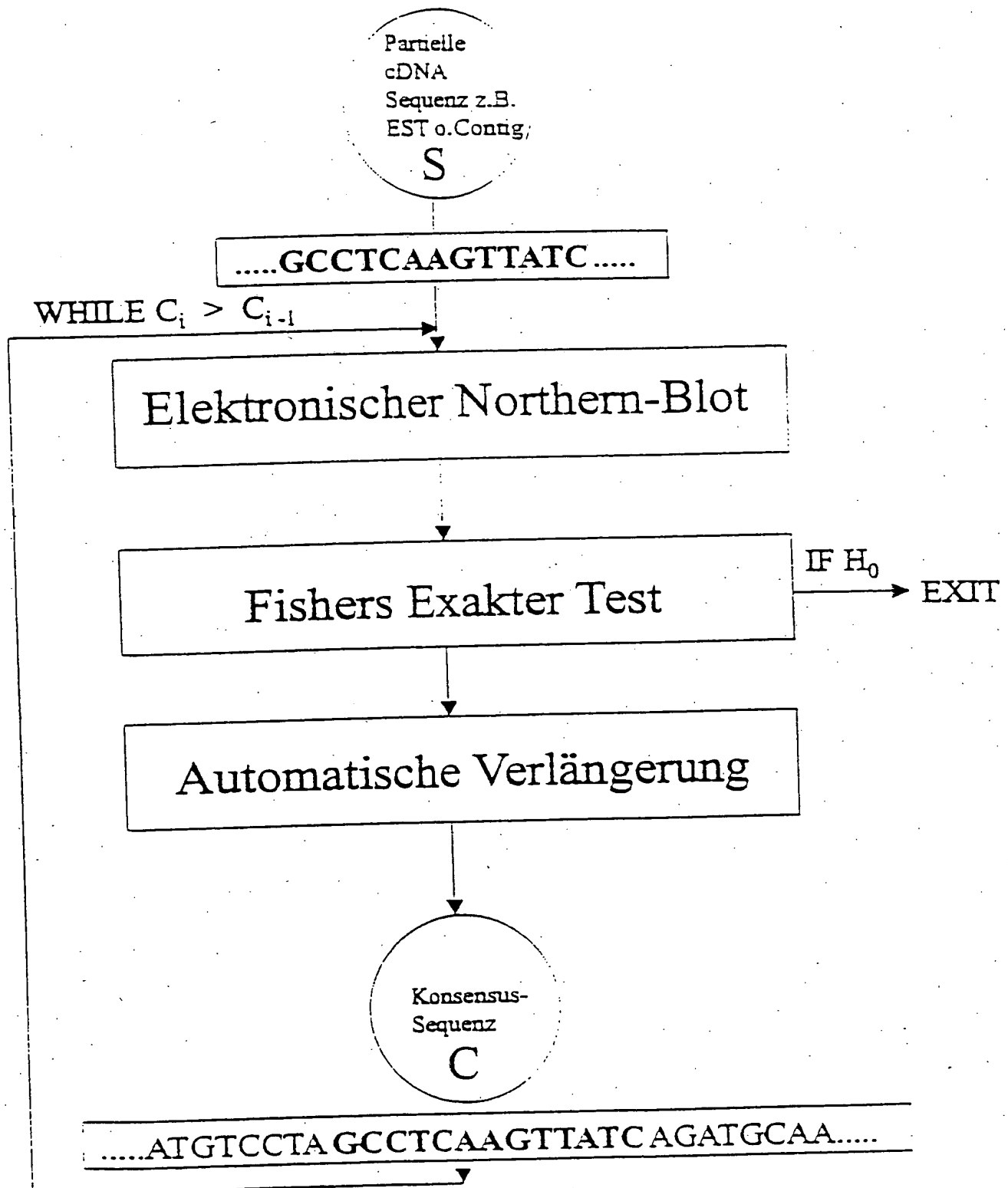


Fig. 4b

10/10

Isolation of genomic BAC and PAC clones

Chromosomal clone localization via FISH

Hybridization signal

Sequencing of clones that are located in regions that have chromosomal deletions in prostate and breast cancer leads to identification of candidate genes

Exon            Intron

Confirmation of candidate genes by screening of mutations and/or deletions in cancer tissues

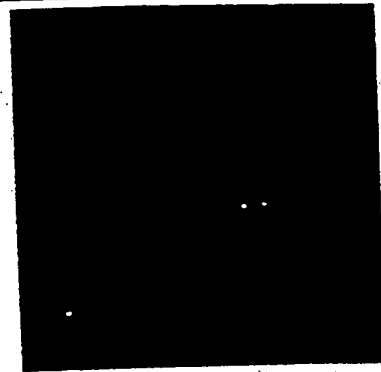
Figure 5

Isolieren von genomischen BAC und PAC Klonen

Chromosomale Klon-Lokalisation über FISH



Hybridisierungssignal



Sequenzierung von Klonen, die in Regionen lokalisiert sind, die chromosomale Deletionen in Prostata- und Brustkrebs aufweisen, führt zur Identifizierung von Kandidatengenen



Bestätigung der Kandidatengene durch Screening von Mutationen und/oder Deletionen in Krebsgeweben

Fig. 5